working paper
department
of economics

FOLK THEOREMS FOR REPEATED GAMES:
a NEU Condition

Dilip Abreu
Prajit K. Dutta
Lones Smith

No.92-15 October 1992

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

FOLK THEOREMS FOR REPEATED GAMES:
a NEU Condition

Dilip Abreu
Prajit K. Dutta
Lones Smith

No.92-15 October 1992

Folk Theorems for Repeated Games: A Neu Condition*

by

Dilip Abreu
Department of Economics
Princeton University
Princeton, NJ 08544

Prajit K. Dutta
Department of Economics
Columbia University
New York, NY 10027

and

Lones Smith
Department of Economics
Massachusetts Institute of Technology
Cambridge, MA 02139

October 1992

*This paper combines "The Folk Theorem for Discounted Repeated Games: A New Condition" by Abreu and Dutta and "Folk Theorems: Two-Dimensionality is (Almost) Enough" by Smith. The pair of papers independently introduced two equivalent conditions, here replaced by a third equivalent condition which is perhaps the most transparent. Abreu and Dutta covered mixed strategies and established the necessity of their condition whereas Smith extended his pure strategy analysis to finitely repeated games and overlapping generation games. We would like to thank David Pearce and Ennio Stacchetti for their comments. The current version reflects helpful financial assistance from the Social Sciences and Humanities Research Council of Canada.

Abstract

The Fudenberg and Maskin folk theorem for discounted repeated games assumes that the set of feasible payoffs is full dimensional. We obtain the same conclusion using a weaker condition. This condition is that no pair of players has equivalent von Neumann-Morgenstern utilities over outcomes. We term this condition NEU ("non-equivalent utilities"). The condition is weak, easily interpreted, and also almost necessary for the result. We also extend our analysis to finitely repeated games and overlapping generations games.

1. Introduction

We are concerned here with "folk theorems" for repeated games with complete information. Such theorems establish that in the limit with little or no discounting any feasible and individually rational payoff of the stage game is an equilibrium payoff of the associated repeated game. More precisely, let $G(A_i, \pi_i : i=1, \ldots, n)$ be a finite (normal form) game, where A_i is a finite action set, M_i is the associated set of mixed actions, and π_i is the payoff function, for player i. Player i's minimax payoff level is denoted $\min_{\mu_{-i}} \max_{a_i} \pi_i(a_i, \mu_{-i})$, and can be normalized to 0 without loss of generality. This is the lowest payoff a maximizing player can be forced down to. It is important that the minimization by the other players be over mixed strategies. Note also the order of the min and max operators: Player i chooses his maximizing action *after* the minimaxing mixed strategy choice of the other players. A payoff vector $\omega = (\omega_1, \ldots, \omega_n)$ is strictly (resp. weakly) *individually rational* if for all $i, \omega_i > (\text{resp. } \geq)\ 0$. Folk theorems assert that *any* feasible and individually rational payoff vector is a (subgame perfect) equilibrium payoff in the associated infinitely repeated game with little or no discounting (where payoff streams are evaluated as average discounted or average values respectively). It is obvious that feasibility and individual rationality are *necessary* conditions for a payoff vector to be an equilibrium payoff. The surprising content of the folk theorems is that these conditions are also (almost) sufficient.

Perhaps the first folk theorem type result is due to Friedman (1971) who showed that any feasible payoff which Pareto dominates a Nash equilibrium payoff of the stage game will be an equilibrium payoff in the associated repeated game with sufficiently patient players. This kind of result is sometimes termed a "Nash threats" folk theorem, a reference to its method of proof. For the more permissive kinds of folk theorems considered here the

seminal results are those of Aumann and Shapley (1976) and Rubinstein (1977, 1979). These authors assume that payoff streams are undiscounted.[1] Fudenberg and Maskin (1986) establish an analogous result for discounted repeated games as the discount factor goes to 1. Their result uses techniques of proof rather different from those used by Aumann-Shapley and Rubinstein, respectively. See their paper for an insightful discussion of this point, and quite generally for more by way of background. It is a key reference for subsequent work in this area, including our own.

For the two-player case, the result of Fudenberg and Maskin (1986) is a complete if and only if characterization (modulo the requirement of strict rather than weak individual rationality, _which we retain in this note_) and does not employ additional conditions. For three or more players Fudenberg and Maskin introduced a _full dimensionality_ condition: the convex hull F, of the set of feasible payoff vectors of the stage game must have dimension n (recall that n is the number of players), or equivalently a non-empty interior. This condition has been widely adopted in proving folk theorems for related environments such as finitely repeated games (Benoit and Krishna (1985)), and overlapping generations games (Smith, (1992)).

Full dimensionality is a _sufficient_ condition. Fudenberg and Maskin present an example of a three-player stage game in which the conclusion of the folk theorem is false. In this example _all_ players receive the same payoffs in all contingencies; the (convex hull of the) set of feasible payoffs is _one_-dimensional. This example violates full dimensionality in a rather extreme way. Less extreme violations may also lead to difficulties as the

[1] Aumann and Shapley (1976) employ the _limit of means_ criterion and Rubinstein (1977) considers both the limit of means and the _overtaking criterion_.

following example taken from Benoit and Krishna (1985) indicates.[2] The payoffs in their three-player example are given by:

3,3,3	0,0,0
0,1,1	0,0,0
0,0,0	0,0,0

1,1,1	2,2,2
0,1,1	0,0,0
0,1,1	0,1,1

where player 1 chooses rows, 2 chooses columns, and 3 chooses matrices. Let $\rho(\delta) = \inf\{\rho | \rho$ is an (average discounted) payoff to players 2 and 3 in some (subgame perfect) equilibrium of the associated repeated game with discount factor $\delta\}$. We argue that for <u>any</u> $\delta > 0$, $\rho(\delta) \geq 1/2$. Since each player's minimax payoff is zero and (1/4, 1/4, 1/4) is, for instance, a feasible payoff, this yields a contradiction to the conclusion of the folk theorem. To establish that $\rho(\delta) \geq 1/2$ we need simply show that for any combination of mixed actions in the stage game <u>either</u> player 2 or 3 can by deviating, if necessary, obtain a payoff of at least 1/2. Suppose this were not the case and let player 1 play row i with probability r_i. Then, if player 2 plays left, he obtains 1 for sure if player 3 chooses the right matrix and $3r_1 + r_2$ otherwise. Hence, $3r_1 + r_2 < 1/2$. Similarly, for player 3 not to receive at least 1/2 by choosing the right matrix, it must be that $2r_1 + r_3 < 1/2$. Together these inequalities yield $4r_1 < 0$, a contradiction.

One might be tempted to conjecture that, except in this special case in which a pair of players have identical payoffs, the conclusion of the folk theorem is true. This guess cannot possibly be exactly right since affine transformations of a player's payoffs do not alter the strategic structure of a game. We must at the very least exclude <u>equivalent</u> von Neumann-Morgenstern

[2] Their analysis was for the finitely repeated case, but the example works equally well in the infinitely repeated setting.

utility functions over outcomes. The functions π_i and π_j are equivalent if there exist scalars c,d where d > 0 such that $\pi_i(a) = c + d\pi_j(a)$ for all $a \in A$. Of course, equivalent utility functions yield identical orderings of lotteries over outcomes. By changing the origin and scale of player i's utility function it may be made identical to player j's. Viewed geometrically, equivalent payoffs lie on a straight line with positive slope.

In this form the initial conjecture is in fact correct; the simple condition that no pair of players have equivalent utility functions is sufficient. We term this requirement non-equivalent utilities (NEU). This condition is easy to understand, and, of course, weaker than full dimensionality. Furthermore, it is a "tight" condition in the sense that it is also often necessary. While full dimensionality may be viewed as a generic condition, NEU holds generically within the smaller class of stage games with dimensions 2,3,...,(n-1) respectively. Such games may arise naturally; for instance the payoffs of various coalitions of players might have constant sum.

Our result is for the standard case in which mixed strategies are unobservable. But to develop some feeling for our condition and the argument assume, for the moment, that mixed strategies are ex-post observable or simply confine attention to pure strategies (and define individual rationality in the latter case through pure strategies). Let u be a strictly individually rational and feasible payoff vector. A simple lemma below shows that an implication of NEU is that there exist n strictly individually rational payoff vectors $x^1,\ldots,x^n$ such that $x_i^i < x_i^j$ $\forall i,j$, $i \neq j$ and $x_i^i < u_i$. We may think of these payoff vectors as player-specific "punishments" which have the property that player i's payoff in his own punishment is strictly worse than his payoff in any other player's punishment. Let a and a^i be (correlated) action profiles which yield the payoff vectors u and x^i, $i=1,\ldots,n$ respectively. Let

P^0 be the path in which a is played in every period and P^i the path in which player i is minimaxed for q periods (during which he plays a best response) followed by the action a^i forever after. Consider the simple profile (Abreu (1988)) in which P^0 is played initially and any deviation by player i alone from an ongoing path is responded to by imposing P^i (and simultaneous deviations are ignored). Let the integer q satisfy $qx^i_i > \max_{a_i} \pi_i(a_i, a^i_{-i}) - \pi_i(a^i)$, for all i. Then, using the criterion of "unimprovability" (which only checks one-shot deviations) it can be directly verified that for high enough discount factors, the described simple profile is a subgame perfect equilibrium. (Player i will not deviate from P^i since by the preceding inequality any one-period gain is wiped out by q periods of minimaxing. Player $j \neq i$ will not deviate from P^i since $x^j_j < x^i_j$. Finally, since $x^i_j < u_i$ a one-shot deviation from P^0 is unprofitable). That is, with observable mixed strategies and given NEU the extension of the undiscounted folk theorem to the discounted case is fairly straightforward. The subtleties in the argument derive primarily from the consideration of mixed strategies.

This paper is organized as follows. Section 2 presents the model. Section 3 establishes the sufficiency of NEU and proves a necessity result in the context of infinitely repeated games with discounting. Section 4 presents analogous results for finitely repeated games. In doing so, we strengthen the Benoit-Krishna (1985) result in two ways: First, we admit the possibility that one player might not have distinct Nash payoffs in the stage game. Second, and this introduces some subtleties that we must finesse in the proof, we produce an exact (rather than approximate) folk theorem. Section 5 discusses overlapping generations games. Section 6 concludes.

2. <u>The Model</u>

We consider a finite n-player game in normal form defined by $<A_i, \pi_i;$ $i=1,\ldots,n>$ where A_i is the i^{th} player's finite set of actions, and let $A = \prod_{i=1}^{n} A_i$. The i^{th} player's payoff is $\pi_i : A \to R$. Let M_i be the set of player i's mixed strategies and let $M = \prod_{i=1}^{n} M_i$. Abusing notation, we write $\pi_i(\mu)$ for i's expected payoff under the mixed strategy $\mu = (\mu_1,\ldots\mu_n) \in M$. For any n-element vector $v = (v_1,\ldots v_n)$ let v_{-i} denote the corresponding (n-1) element vector with the i^{th} element missing. Let $\pi^*_i(\mu_{-i}) = \max_{a_i} \pi_i(a_i,\mu_{-i})$ be player i's best response payoff against the mixed profile μ_{-i}. Denote by $m^i =$ $(m^i_1,\ldots,m^i_n) \in M$ a mixed strategy profile which satisfies $m^i_{-i} \in$ $\operatorname{argmin}_{\mu_{-i}} \pi^*_i(\mu_{-i})$ and $m^i_i \in \operatorname{argmax}_{\mu_i} \pi_i(\mu_i, m^i_{-i})$ In words, m^i_{-i} is a (n-1)-profile of mixed strategies which minimax player i and m^i_i is a best response for i when being minimaxed. We have adopted the normalization $\pi_i(m^i) = 0$ for all i. Let $F = \operatorname{co}\{\pi(\mu):\mu \in M\}$ and denote by F* the set of feasible and (strictly) individually rational payoffs; $F^* = \{w \in F: w_i > 0, \text{ for all } i\}$.

We will analyze the associated repeated game with perfect monitoring. That is, for all i, player i's action in period t can be conditioned on the past actions of all players. In addition, we permit <u>public randomization</u>. That is, in every period players publicly observe the realization of an exogeneous continuous random variable and can condition on its outcome. This assumption can be made without loss of generality; a result due to Fudenberg and Maskin (1991) shows explicitly how public randomization can be replaced by "time-averaging" in the infinitely repeated case.

Denote by $\alpha_i = (\alpha_{i1},\ldots,\alpha_{it},\ldots)$ a (behavior) strategy for player i in the repeated game and by $\pi_{it}(\alpha)$ his expected payoff in period t given the

strategy profile α. Player i's (average) discounted payoff under the (common) discount factor δ is $v_i(\alpha) = (1-\delta)\sum_0^\infty \delta^t \pi_{it}(\alpha)$. Let $V(\delta)$ denote the set of subgame perfect equilibrium payoffs.

3. Infinitely Repeated Games

NEU has two quite powerful and equivalent representations, one of which has already been alluded to in the informal discussion above. These are developed in the lemmas below.

Let F_{ij}, $j \neq i$ denote the projection of F on the i-j coordinate plane and dim F_{ij} the dimension of F_{ij}. Observe that if F^*, the set of feasible and strictly individually rational payoffs, is non-null, then no player is indifferent over all possible action profiles.

<u>Lemma 1</u> Suppose NEU and that no player is indifferent over all possible profiles. Then the following projection condition obtains: for all $i \neq j$, dim $F_{ij} \geq 1$; furthermore, if for some $j \neq i$, dim $F_{ij} = 1$, then F_{ij} has a strictly negative slope and dim $F_{ik} = 2$ for all $k \neq i,j$.

<u>Proof</u> Since we rule out universal indifference by assumption, it is clear that dim $F_{ij} \geq 1$ for all i,j, $i \neq j$. Now suppose that $\dim(F_{ij}) = \dim(F_{ik}) = 1$ for some $j \neq k$. NEU applied to the payoffs of players i and j (and similarly players i and k) implies that the payoffs are perfectly negatively correlated. This in turn implies that the payoffs of players j and k are perfectly positively correlated. That is, players j and k have equivalent payoffs, in violation of NEU. Q.E.D.

The representation below was already discussed in the introduction.

__Lemma 2__ Suppose that the projection condition of Lemma 1 obtains. Let u be a feasible and strictly individually rational payoff vector. Then there exist payoff vectors $x^i, i=1,\ldots,n$ such that $\forall i,j, i\neq j$,

1. $x^i \gg 0$ strict individual rationality
2. $x^i_i < x^j_i$ payoff asymmetry
3. $x^i_i < u_i$ target payoff domination

__Step 1__ By Lemma 1 it follows that for all $i,j,i\neq j$ there exist feasible payoff vectors v^{ij} such that $v^{ij}_i > v^{ji}_i$ and $v^{ij}_j < v^{ji}_j$. Let θ_k, $k=1,\ldots,\frac{n(n-1)}{2}$ be strictly increasing weights (with $\sum \theta_k = 1$) and v^i be a convex of all the v^{kj}'swhich gives weight θ_k to player i's k^{th} best payoff (break ties arbitrarily) in the set $\{v^{kj}\}_{k\neq j}$. Obviously $v^i_i < v^j_i$ for all $i,j,i\neq j$.

__Step 2__ Let w^i denote a payoff vector which yields player i his lowest payoff in the game; i.e., $w^i_i = \min\{v_i:(v_{-i},v_i) \in F\}$. Define

$$x^i = \beta_1 w^i + \beta_2 v^i + \beta_3 u$$

where β_1,β_2,β_3 are convexifying weights __which are independent of i__ and chosen such that β_2 is strictly positive and the ratios β_2/β_1 and $(\beta_1+\beta_2)/\beta_3$ are both small. By the definition of w^i and v^i it follows that if β_2 is strictly positive, $x^i_i < x^j_i$ for all $i,j,i \neq j$ (payoff asymmetry). For small enough β_2/β_1, we must have $x^i_i < u_i$ (even if $v^i_i > v_i$). Thus we can obtain target payoff domination. Finally for small enough $(\beta_1+\beta_2)/\beta_3$, $x^i_i > 0$ since $u_i > 0$ (strict individual rationality). Q.E.D.

Lemma 2 also has an easy geometric proof. For consider u as a point in R^n, and let Π be the vector space spanned by the payoff space F. Note that Π might be high dimensional and hard to visualize. With a little imagination, however, it can be seen that we may choose some (two-dimensional) plane Π^* within Π satisfying the projection condition of Lemma 1. That is, the

projection of Π^* onto any two players' coordinate plane is either two-dimensional or a line with negative slope. We may now select our vectors $\{x^i\}$ to lie on Π^*. Here's how: Draw an $\epsilon > 0$ circle about u in Π^*. For sufficiently small $\epsilon > 0$, all points on it are strictly individually rational. For each i, let x^i be the point on the circle with the smallest i-coordinate. We need only show that the $\{x^i\}$ are distinct. By assumption, the projection of the circle onto the coordinate plane of any two players (i,k) is either a line of negative slope or an ellipse. In the first case, x^i and x^k lie at opposite ends of a line segment, while in the second, they lie at different locations on the ellipse.

It is straightforward to see that the existence of asymmetric payoff vectors x^i (satisfying inequality 2. above) implies NEU. In other words, NEU, the projection condition and the existence of asymmetric payoffs are equivalent assumptions (modulo $F^* \neq \phi$).

We are now ready to prove the folk theorem.

Theorem 1 Under NEU, any point in F^* is a subgame perfect equilibrium payoff when players are sufficiently patient. That is, for any $u \in F^*$ there exists $\underline{\delta} < 1$ such that $u \in V(\delta)$ for all $\delta \geq \underline{\delta}$.

Proof Fix $u \in F^*$ and let $x^i, i=1,\dots,n$ be as defined in Lemma 2.

We will assume without loss of generality that, in addition, the x^i's lie in the relative interior of F^*. We now specify a strategy profile which yields payoff u and which for sufficiently high δ is also a subgame perfect equilibrium. The strategy is as follows. Players play the (correlated) action that generates payoff u at t=1 and continue to do so unless some player i deviates singly in some period t. The key element of the specification is the "punishment" for player i which is invoked for any (single person)

deviation by player i from prescribed behavior. Assume that for some $k \neq i$, $\dim(F_{ik}) = 1$. (The proof when there is no such k is a corollary). Then by Lemma 1, $\dim(F_{ij}) = 2$, for $j \neq i$, k. Furthermore, observe that by NEU, F_{ik} is a straight line with negative slope. Let m^i minimax player i. Then fixing m^i_j, for $j \neq i$, k, induces a constant-sum game between i and k. We will require that player i's best response m^i_i additionally minimaxes player k in this induced constant-sum game. Lemma 1 and this observation are the key new elements of our proof.

The punishment for player i consists of q periods of play of m^i as specified above followed by transition to action profiles which yield player i, x^i_i. The difficulty now is to induce minimaxing players $j \neq i,k$ to play pure strategies in the support of their mixed strategies with the appropriate probabilities. As noted by Fudenberg and Maskin (1986), the only way to do so is to make them indifferent across their pure strategies. Note that player k's mixed strategy m^i_k is a best response to m^i_{-k}, and that for $j \neq i,k$ deviations outside the support of m^i_j are easily deterred by directly punishing player j.

For $j \neq i,k$, since $\dim(F_{ij}) = 2$ and the x^i's are in the relative interior of F^*, there exist payoff vectors $c^{ij} \in F^*$ such that

Asymmetry $\qquad x^{\ell}_{\ell} < c^{ij}_{\ell}, \ \ell \neq i \qquad$ (3.1)

Indifference for i $\qquad x^i_i = c^{ij}_i$

Differential for j $\qquad x^i_j > c^{ij}_j \qquad$ (3.2)

At the end of q periods of minimaxing player i, play moves probabilistically (via the public randomization device) to x^i or to the c^{ij}'s, where the probabilities are chosen to maintain the incentives of minimaxing players $j \neq i$. Let $p^{ij}_t(a_j)$ satisfy for all a_j, a'_j in the support of m^i_j, and $t \leq q$,

$$(1-\delta)\pi_j(a_j, m^i_{-j}) + \delta^{q-t+1}\, p^{ij}_t(a_j)(c^{ij}_j - x^i_j) =$$

$$(1-\delta)\pi_j(a'_j, m^i_{-j}) + \delta^{q-t+1}\, p^{ij}_t(a'_j)(c^{ij}_j - x^i_j) \qquad (3.3)$$

The values $p^{ij}(a_j)$ are not unique. One procedure through which they can be defined is as follows: suppose that action a_j yields the highest myopic payoff to player j (when the other players play m^i_{-j}). Then set $p^{ij}_t(a_j)$ equal to 0. For other actions, a'_j, $p^{ij}_t(a'_j)$ is defined through (3.3) after substituting $p^{ij}_t(a_j) = 0$.

If the realized sequence of action profiles is a_t, $t=1,\ldots,q$ then for all $j\neq i,k$ play proceeds to c^{ij} with probability $\sum_1^q p^{ij}_t(a_{jt})$. Notice that the p^{ij}_t's are independent across players $j\neq i$ and across periods and are chosen to make players indifferent across the support of their minimaxing strategies. Let $p^{ij}(a_1,\ldots,a_q) = \sum_1^q p^{ij}_t(a_{jt})$. Then with probability $1 - \sum_j p^{ij}(a_1\ldots,a_q)$, play goes to x^i. For high enough δ, the p^{ij}'s defined above can be made small and positive for all possible realizations of $a_1,\ldots,a_q$ and so we indeed have probabilistic transitions. Since the inequalities, (3.1) and (3.2) are strict, no detectable one-shot deviation is profitable. By construction, minimaxing players are indifferent over the support of their mixed minimaxing strategies. It follows that for the profile specified, after no history does a one-shot deviation yield a higher payoff. By the "unimprovability" criterion, the profile is a subgame perfect equilibrium. Q.E.D.

We turn now to the <u>necessity</u> of NEU. To avoid trivialities, from here on $F^* \neq \phi$. Let $f_i = \min\{v_i | v \in F \text{ and } v_j \geq 0 \text{ for all } j\}$. Thus f_i is the worst payoff to i in the set of weakly individually rational payoff <u>vectors</u>. We will refer to f_i as play i's <u>minimal attainable payoff</u>. <u>The necessity</u> of payoff asymmetry is shown for games in which no two (maximizing) players can

be simultaneously held at or below their minimal attainable payoff. In stage games where every player's minimal attainable payoff is indeed his minimax payoff ($f_i = 0$), the condition stated below is equivalent to the restriction that no pair of players can be simultaneously minimaxed.[3] Note that the condition below uses the term mini<u>miz</u>ing.

<u>No simultaneous Minimizing</u> [NSM] For all $\mu \in M$ such that $\pi^*_i(\mu_{-i}) \leq f_i$ for some i, it is the case that $\pi^*_j(\mu_{-j}) > f_j$ for all $j \neq i$.

Under the above assumption we obtain a <u>complete</u> characterization.

<u>Theorem 2</u> Suppose NSM obtains. Then, NEU is both necessary and sufficient for the conclusion of the folk theorem for discounted infinitely repeated games.

<u>Proof</u> To establish necessity, we exhibit payoff vectors w^i $i=1,\ldots,n$ such that for all $i,j,i \neq j$, $w^i_i < w^j_i$. It follows immediately that for all i,j, $i \neq j$ the outcome corresponding to w^i is strictly worse for player i than the outcome corresponding to w^j, and conversely for player j, so that NEU is satisfied.

Since the conclusion of the folk theorem is valid, the set of subgame perfect equilibrium payoffs is non-empty for sufficiently high δ. Denote by $w^i(\delta)$ an equilibrium payoff vector which yields player i his lowest subgame perfect equilibrium payoff. By adapting the argument of Abreu, Pearce and Stacchetti (1990) it can easily be shown that $w^i(\delta)$ exists;[4] denote by α^i an

---

[3] A game in which the minimally attainable payoff is not the minimax payoff for every player is the two-player game specified by the following payoffs: co[(0,-1),(1,0),(2,1)] with (0,-1) a payoff at which player 1 is minimaxed and (1,0) a payoff where player 2 is minimaxed.

[4] The available results on the existence of the worst equilibrium are for the case where mixed strategies are observable; hence we cannot directly appeal to any of them. The result is however certainly true and may be proved by, for instance, adapting the self-generation techniques of Abreu, Pearce and

equilibrium strategy profile that generates $w^i(\delta)$. By definition, $w^j_j(\delta) \leq w^i_j(\delta)$ for all i,j. By the folk theorem hypothesis, $w^i_i(\delta) \to f_i$, as $\delta \to 1$. By playing his myopic best response in period one and conforming thereafter, player i can get at least $(1-\delta)\ \pi^*_i(\gamma^i_{-i}(\delta)) + \delta w^i_i(\delta)$ where $\gamma^i(\delta)$ is the first period strategy vector in the play of α^i. Hence, $w^i_i(\delta) \geq (1-\delta)\pi^*_i(\gamma^i_{-i}(\delta)) + \delta w^i_i(\delta)$, or equivalently $w^i_i(\delta) \geq \pi^*_i(\gamma^i_{-i}(\delta))$. It then follows that as $\delta \to 1$ any subsequential limit of $\gamma^i_{-i}(\delta)$ yields player i, in a best response, a payoff at or below f_i, i.e. $\pi^*_i\ (\lim\gamma^i_{-i}(\delta)) \leq f_i$.

Clearly $w^i_j(\delta) \geq (1-\delta)\ \pi^*_j(\gamma^i_{-j}(\delta)) + \delta w^j_j(\delta)$. Hence if $w^i_j(\delta) = w^j_j(\delta)$ it follows that $w^i_j(\delta) \geq \pi^*_j(\gamma^i_{-j}(\delta))$. We claim now that there is $\underline{\delta} < 1$ such that $w^j_j(\underline{\delta}) < w^i_j(\underline{\delta})$ for all i,j, j≠i. A contradiction to this claim implies the existence of a sequence $\delta_m \to 1$ and fixed indices i,j, j≠i such that $w^j_j(\delta_m) \geq \pi^*_j(\gamma^i_{-j}(\delta_m))$. The left-hand side of the inequality goes to f_j while the right-hand side is strictly greater than f_j since $\pi^*_i(\lim\gamma^i_{-i}(\delta)) \leq f_i$ and simultaneous minimizing is impossible (NSM). This yields the desired contradiction. Now take $w^i = w^i(\underline{\delta})$. Q.E.D.

Remarks:

1. When NSM is not satisfied a weaker version of NEU is necessary for the folk theorem. This condition is that any subset of players who cannot all be simultaneously minimized cannot all have equivalent utilities. It must be possible to simultaneously minimize the entire subset of players who have equivalent payoffs.

2. One class of games in which NSM is violated is two-player games. Another is the class of symmetric games in which all players can be simultaneously minimaxed. By the preceding remark all players may have equivalent utilities

---

Stachetti (1990) to the present context.

in such games.[5,6] Indeed, for these games it essentially follows from the two-player analysis of Fudenberg and Maskin (1986) that the folk theorem holds without any conditions on the set of feasible payoffs.

4. The Folk Theorem for Finitely-Repeated Games

In this section, we let $G(\delta,T)$ denote the T-fold repetition of G with the discount factor $\delta \leq 1$. What follows is a strengthening of Theorem 3.7 of Benoit and Krishna (1985), (or the simpler proof in Krishna (1989)) on several fronts. First, we admit payoff discounting. In so doing, we can establish a uniform folk theorem, meaning that the discount factor and horizon length can vary independently over the relevant range. Second, we substitute NEU for full-dimensionality. Finally, we weaken the condition that all players must have distinct Nash payoffs to all but one -- provided his Nash payoff is strictly positive. However, unlike the (more intuitive) use of long deterministic cycles in Benoit and Krishna (1985), we simply rely upon correlated outcomes. This has the useful biproduct of permitting an exact (rather than approximate) folk theorem.

Let the best and worst Nash payoff vectors for player i in the stage game G be y^i and z^i, $\forall i$. Also, denote the set of subgame perfect equilibrium payoffs by $V(\delta,T)$.

Theorem 3

Suppose G satisfies NEU, and that either (a) every player has a strict ranking among some pair of Nash outcomes, i.e. $y^k_k > z^k_k$ $\forall k$, or (b) all but player 1 does, and that player 1's unique Nash payoff is strictly individually

---

[5] For symmetric games, $f_i = f_j = 0$. Hence NSM reduces precisely to no simultaneous minimaxing.

[6] In any symmetric game it is always possible to simultaneously minimax two players, but not necessarily all players.

rational, i.e. $y^k_k > z^k_k$ $\forall$ $k \neq 1$, but $y^1_1 = z^1_1 > 0$. Then any point in F^* is a subgame perfect equilibrium payoff of the finitely-repeated game when players are sufficiently patient and the horizon is long enough. That is: $\forall$ $u \in F^*$, $\exists$ $T_0 < \infty$ and $\delta_0 < 1$ so that $T \geq T_0$ and $\delta \in [\delta_0, 1] \Rightarrow u \in V(\delta, T)$.

<u>Proof</u>

We suppose first that mixed strategies are observable, and discuss how to modify the argument later. Let y^* be a payoff vector according equal weight $1/n$ (via public randomizations) to each of the <u>preferred</u> Nash payoff vectors y^i. The typical T-period equilibrium outcome sequence is $\tilde{u}, \ldots, \tilde{u}$; $y^*, \ldots, y^*$, where y^* lasts $s < T$ periods, and $\tilde{u} \in F^*$ will be seen to satisfy the required target payoff equation

$$(T-s)\tilde{u} + sy^* = Tu \tag{4.1a}$$

if $\delta = 1$, or

$$(1-\delta^{T-s})\tilde{u} + \delta^{T-s}(1-\delta^s)y^* = (1-\delta^T)u \tag{4.1b}$$

if $\delta < 1$. We now explicitly describe players' strategies. For ease of exposition, <u>late</u> deviations are those occurring during the final $q + r + s$ periods of the repeated game; all others are called <u>early</u> deviations.[7] Also note that we interchangeably refer to an action and its associated payoff vector.

1. Play $\tilde{u}$ until period $T - s$. [If player j deviates early, start 2; if player 1 deviates late, start 4.] Then play y^* until the end.

2. Play m^j for q periods. [If player $k \neq j$ deviates, start 3.] Then set $k \leftarrow j$.

[7] In this proof, j,k, and ℓ denote arbitrary players. Moreover, for clarity, we use the simple notation $k \leftarrow j$ to mean "assign k the value j." Also, program steps always follow sequentially, unless otherwise indicated.

3. Play x^k for r periods. [If some player j deviates early, restart 2; if player $\ell \neq 1$ deviates late, start 4; if 1 deviates late, start 5.] Then return to step 1.

4. Play z^ℓ until the end.

5. Play m^1 until period $T - s + \lfloor\sqrt{s}\rfloor$ [If player $\ell \neq 1$ deviates, start 4.] Then play y^* until the end.

Notice that step 5 reflects our weakening of the distinct Nash payoff requirement. Effectively, we enlist the support of players 2,3,...,n (who do have distinct Nash payoffs) to ensure <u>late</u> compliance by player 1 (who might not).

We proceed recursively as we ensure subgame perfection. First note that given continuity of discounted sums in δ, if each deterrent is strict by some positive margin, say 1, they will remain strict for any level of discounting $\delta \in [\delta_0, 1]$, for some $\delta_0 < 1$.

We now choose q, r, and s. In light of $x_j^j > 0$, let q satisfy[8]

$$w_j^j + qx_j^j > b_j^j + 1 \tag{4.2}$$

for all $j \in N$, where b^j is the best payoff vector for player j in G. Since $x_j^i < u_j$ too, (4.2) simultaneously renders the punishment interlude a strict deterrent to deviations from steps 1 and 3, for any r. Next, step 3 deters deviations by the punishers from step 2 if r is large enough that

$$qw_k^k + rx_k^j > b_k^k + rx_k^k + (q - 1)\tilde{u}_k + 1 \tag{4.3}$$

[8] The inequality (4.2) -- in particular, why the left-hand side is <u>not</u> $(q+1)x_j^i$ -- reflects the fact that obeying the random correlating device generating x^i might sometimes require j to play his worst possible outcome.

for all $k \neq j$. Given the $\tilde{u}$ is as yet unspecified, we shall instead insist that r satisfy

$$qw_k^k + rx_k^j > b_k^k + rx_k^k + (q - 1)(2u_k - x_k^k) + 1. \tag{4.4}$$

This will turn out to imply (4.3) once $\tilde{u}$ is chosen. Thus, the punishment interlude, steps 2 and 3, will deter deviations from step 1.

Next step 4 will deter all "late" deviations by players $\ell \neq 1$ so long as s is large enough that

$$(q+r+\sqrt{s})w_\ell^\ell + (s-\sqrt{s})y_\ell^* > b_\ell^\ell + (q+r+s-1)z_\ell^\ell + 1 \tag{4.5a}$$

for all $\ell \neq 1$. This inequality obtains for all sufficiently large s, because for each player $y_\ell^* > z_\ell^\ell$ for all $\ell \neq 1$, and since s grows faster than $\sqrt{s}$. Similarly, step 5 is a deterrent to late deviations by player 1 from steps 3 and 5, respectively, provided

$$qw_1^1 + rx_1^1 + sy_1^* > b_1^1 + (s - \sqrt{s})y_1^* + 1 \tag{4.5b}$$

and

$$sy_1^* > b_1^1 + (s-\sqrt{s})y_1^* + 1. \tag{4.5c}$$

Clearly, for large enough s, inequalities (4.5a), (4.5b), and (4.5c) are valid.

Recall that the personalized punishment vectors satisfy $x_k^k < u_k$ $\forall k$. It now remains to ensure that T is always large enough that (1) yields $\tilde{u}$ close enough to u so that $x_k^k < \tilde{u}_k$ $\forall k$. We continue along these lines: We first specify a lower bound T_0 for T, and then proceed to verify that for any $\delta \geq \delta_0$ (appropriately chosen), the implied $\tilde{u}$ works.

Given q, r, and s as defined above, feasibility entails $T_0 \geq q + r + s$. Next, since $u \in F^*$, there is some η-neighborhood (in the hyperplane which F^*

spans) around u entirely contained within F^*. Then let T_0 be sufficiently large that

$$\frac{s}{T_0 - s} \| u-y^* \| < \min(\eta, \min_j [u_j - x_j^j]) \tag{4.6a}$$

so that

$$\delta^{T-s} \frac{1 - \delta^s}{1 - \delta^{T-s}} \| u-y^* \| < \min(\eta, \min_j [u_j - x_j^j]) \tag{4.6b}$$

for all $\delta \leq 1$ and $T \geq T_0$.

Finally, given $T \geq T_0$, let each deterrent remain strict for all $\delta \in [\delta_o, 1]$, for some $\delta_o < 1$. Define $\tilde{u}$ implicitly by the target payoff equation (1a) or (1b). To verify that indeed $\tilde{u} \in F^*$, (4.1a) and (4.1b) can be rewritten as

$$\tilde{u} - u = \begin{cases} \frac{s}{T-s} [u-y^*] & \text{if } \delta = 1 \\ \delta^{T-s} \frac{1-\delta^s}{1-\delta^{T-s}} [u-y^*] & \text{if } \delta < 1 \end{cases} \tag{4.7}$$

In light of (4.6a) and (4.6b), equation (4.7) both says that $\tilde{u} \in F^*$ and also that

$$| \tilde{u}_k - u_k | < \min_j [u_j - x_j^j] \tag{4.8}$$

for all k. By the triangle inequality, there are two immediate consequences of (4.8). First, we may conclude that $\tilde{u}_k < 2u_k - x_k^k$, so that (4.4) implies (4.3); and second, that

$$\begin{aligned} \tilde{u}_k - x_k^k &= u_k - x_k^k + \tilde{u}_k - u_k \\ &\geq [u_k - x_k^k] - |\tilde{u}_k - u_k| > 0 \end{aligned}$$

for all k, both of which were asserted earlier. Q.E.D.

We now address the knotty issue of unobservable mixed strategies. Suppose that δ has been chosen. Define c^{ij} just as in (4.6), but now let $p_t^{ij}(a_j)$ satisfy for all a_j, a'_j in the support of m_j^i, and $t \le q$,

$$(1-\delta)\pi_j(a_j, m_{-j}^i) + [\delta^{q-t+1} - \delta^{q+r-t+1}]p_t^{ij}(a_j)(c_j^{ij} - x_j^i) =$$

$$(1-\delta)\pi_j(a'_j, m_{-j}^i) + [\delta^{q-t+1} - \delta^{q+r-t+1}]p_t^{ij}(a'_j)(c_j^{ij} - x_j^i).$$

Thus we may modify, à la Theorem 1, our step 3. That is, play proceeds probabilistically to one of several s-period phases, <u>not</u> an infinitely long phase.

Consider now the necessity of the conditions of Theorem 3. We do not address the necessity of the proviso about distinct Nash equilibrium payoffs. Rather, we consider the class $\mathcal{G}$ of games satisfying that criterion, and ask whether NEU is needed. Suppose not. To wit, let a folk theorem obtain when the one-shot game is G^*. Now imagine the supergame with stage game G^* as a sequence of games $G^*(\delta,T)$. Clearly, if u is a subgame perfect discounted average payoff for $G^*(\delta,T)$, then it is also one for the infinitely repeated game with stage game G^*. This establishes

<u>Theorem 4</u>

Suppose that NSM obtains. Then NEU is both necessary and sufficient for the conclusion of the finite horizon folk theorem in the class of games $\mathcal{G}$.

5. A Folk Theorem for Overlapping Generations Games

Smith (1992) considers a model of overlapping generations games OLG(G;δ,T), in which -- in its most basic formulation -- a player dies and is

replaced every T periods.[9] Using the techniques developed in the previous section, it is possible to render that folk theorem exact too. Moreover, its full-dimensionality condition can also be weakened to payoff asymmetry.

Theorem 5

Let $u \in F^*$. Suppose that NEU holds. Then any point in F^* is a subgame perfect equilibrium payoff of the overlapping generations game when players are sufficiently patient and the overlap between player deaths is long enough. That is, for any $u \in F^*$, $\exists\, T_0 < \infty$ and $\delta_0 < 1$ so that $T \geq T_0$ and $\delta \in [\delta_0, 1] \Rightarrow$ OLG(G;δ,T) has a subgame perfect discounted average payoff u.

A parallel proof of this result is possible, but is omitted.

Once again, we ask if NEU is necessary here. The answer is not as immediate. For in an overlapping generations context, because the players' tenures do not coincide, we may also distinguish between them <u>intertemporally</u>. That is, we may await the death of specific players before rewarding or punishing the others. The development of this idea is the substance of the <u>non-uniform</u> folk theorems of Smith (1992) and Kandori (1992). Indeed, such a folk theorem can obtain <u>even when all players receive the same payoff</u>; however, any non-uniform result demands that the discount factor and horizon length <u>positively covary</u>. That is, $\delta \uparrow 1$ necessarily as $T \uparrow \infty$.

This insight allows us to more generally think of NEU more generally as just one method of awarding players separable payoff streams. For the uniform result of Theorem 5, we must payoff distinguish between players at the stage game level; punishments cannot be deferred. We must leave as an open question the necessity of NEU for a uniform folk theorem for an overlapping generations game. Our strong suspicion is that it is necessary, and that a convex

---

[9] Kandori (1992) describes a related model, but doesn't consider the uniform folk theorem discussed below.

combination of the methods of Theorem 2 and of example 3 in Smith (1992) will eventually do the trick.

6. Concluding Remarks

We have established folk theorems by assuming that players have non-equivalent utilities. This condition is weaker than the full dimensionality condition introduced by Fudenberg and Maskin (1986). Our condition is appealing in that it is easily interpreted (while full dimensionality is a natural geometric concept, it lacks an immediate strategic interpretation) and also minimal in the sense of being almost necessary. It focuses on the deterrence of individual deviations as required by Nash equilibrium theory; full dimensionality permits the greater but unnecessary luxury of providing individually calibrated punishments to all players simultaneously.

Dutta (1991) uses some of the ideas presented here in proving a folk theorem for the more general class of stochastic games. An interesting question is whether our results can be extended to other environments, such as imperfect monitoring, in which full dimensionality has been invoked (see Fudenberg, Levine, and Maskin (1989)) to prove folk theorems.

References

Abreu, D., 1988, On the Theory of Infinitely Repeated Games with Discounting, Econometrica 56, 383-396.

Abreu, D., D. Pearce, and E. Stacchetti, 1990, Towards a Theory of Discounted Repeated Games with Imperfect Monitoring, Econometrica 58, 1041-1063.

Abreu, D. and P. Dutta, 1991, The Folk Theorem for Discounted Repeated Games: A New Condition, mimeo.

Aumann, R. and L. Shapley, 1976, Long Term Competition: A Game Theoretic Analysis, mimeo, Hebrew University.

Benoit J.-P. and V. Krishna, 1985, Finitely Repeated Games, Econometrica 53, 905-922.

Dutta, P.K., 1991, A Folk Theorem for Stochastic Games, University of Rochester Working paper # 293.

Friedman, J., 1971, A Noncooperative Equilibrium for Supergames, Review of Economic Studies 38, 1-12.

Fudenberg, D., D. Levine and E. Maskin, 1989, The Folk Theorem with Imperfect Public Information, mimeo, UCLA.

Fudenberg, D. and E. Maskin, 1986, The Folk Theorem in Repeated Games with Discounting or Incomplete Information, Econometrica 54, 533-554.

Fudenberg, D. and E. Maskin, 1991, On the Dispensability of Public Randomization in Discounted Repeated Games, Journal of Economic Theory 53, 428-438.

Kandori, M., 1992, Repeated Games Played by Overlapping Generations of Players, Review of Economic Studies 59, 81-92.

Krishna, 1989, The Folk Theorems for Repeated Games, mimeo 89-003, Harvard Business School.

Rubinstein, A., 1977, Equilibrium in Supergames, Research Memorandum #25. Center for Research in Mathematical Economics and Game Theory, Hebrew University.

Rubinstein, A., 1979, Equilibrium in Supergames with the Overtaking Criterion, Journal of Economic Theory 31, 227-250.

Smith, L., 1990/1992, Folk Theorems: Two-Dimensionality is (Almost) Enough, mimeo, MIT.

Smith, L., 1992, Folk Theorems in Overlapping Generations Games, Games and Economic Behavior 4, 426-449.